Journal of Forensic and Legal Medicine 17 (2010) 67–77

Contents lists available at ScienceDirect

Journal of Forensic and Legal Medicine

journal homepage: www.elsevier.com/jflm

Review

Applications and limitations of Forensic Biomechanics: A Bayesian perspective

Michael D. Freeman PhD MPH (Clinical Associate Professor of Epidemiology, Adjunct Associate Professor of Forensic Medicine and Epidemiology)[a,b,*], Sean S. Kohles PhD (Research Associate Professor of Mechanical & Materials Engineering, Adjunct Associate Professor of Surgery)[c,d]

[a] Department of Public Health and Preventive Medicine, Oregon Health and Science University School of Medicine, United States
[b] Institute of Forensic Medicine, Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark
[c] Department of Mechanical and Materials Engineering, Portland State University, United States
[d] Department of Surgery, Oregon Health and Science University School of Medicine, United States

ARTICLE INFO

Article history:
Received 24 January 2009
Received in revised form 7 July 2009
Accepted 9 September 2009
Available online 9 October 2009

Keywords:
Forensic Biomechanics
Forensic epidemiology
Bayes' Law
Error Odds

ABSTRACT

Forensic Biomechanics is an analytic method intended for presentation in a court of law. The method consists of the reconstruction of an injury mechanism followed by a comparison between the injury risk of the mechanism and the injury tolerance of the individual. In recent years some courts have excluded such testimony based, in part, on the inability of experts to quantify the potential error of the methods they relied upon in reaching their conclusions. The application of Bayes' Law to a forensic test of truth in a disputed matter allows for quantification of the error inherent in the method through the conditioning of the pre-test probability of the test outcome with the true and false positive rate of the test. The result of the calculation is the Error Odds (O_E) for the test, or the ratio of correct to incorrect tests.

We present an Error Odds analysis of seven previously published case studies in Forensic Biomechanics as an illustration of the utility of the O_E as a metric for admissibility of testimony in the courts, with a minimum Error Odds ratio of 10 proposed as a threshold. The results of our analysis yielded only 1 of 7 cases of applied Forensic Biomechanics that surpassed the threshold for admissible testimony of 10, with most the cases falling below an O_E of 3. The results of the present study suggest that the forensic application of biomechanics is potentially fraught with error. We suggest that an Error Odds analysis be incorporated in Forensic Biomechanics as part of the analysis as a form of quality control and as demonstrable evidence of the accuracy of the methodology.

1. Introduction

In their paper entitled ‘‘Forensic Injury Biomechanics,’’ Hayes et al. describe a method for estimating injury probability as a means of establishing the ‘‘biomechanical plausibility’’ of an injury when there is a dispute over the cause or some other aspect of the injury.[1] These authors advocate a three step method of biomechanical injury causation primarily directed at traffic collision injuries that first reconstructs the injury event forces, then uses a computer model of an anthropometric test dummy (ATD) to recreate the occupant movement and forces, and finally assigns a Factor of Risk (Φ) that represents the ratio of the crash forces to a previously described injury tolerance metric. The results are used most often in civil legal settings to either explain how an injury occurred (when used for the plaintiff), assert that an injury was unlikely to have occurred (when used for the defense), or describe the effect on injury risk of a hypothetical situation, such as the use of a seat belt (for either plaintiff or defendant).

These authors propose an intriguing concept; a biomechanical test for truth in disputed matters. The proposal is new territory for applied biomechanics; prior use of biomechanical analysis of real world (vs. experimentally produced) injuries has been largely confined to traffic crashes, and associated with efforts by the National Highway Traffic Safety Administration (NHTSA) or the National Transportation Safety Board (NTSB) to analyze crash-related injuries as a means of explaining them, or evaluating the effect that a safety device like a seat belt or airbag may have had on the outcome.[2,3] The initial determination of injury causation made by clinicians for NHTSA and NTSB (i.e. that the injury resulted from the investigated collision) is not altered by the biomechanical investigation; it is explicated. This differs fundamentally from the system described by Hayes et al. which is proposed as a means of determining ‘‘whether’’ rather than ‘‘how’’, for example, a herniated disc in the neck could result from a low-speed rear-end collision, among other tasks.

* Corresponding author. Address: 205 Liberty Street NE, Suite B, Salem, OR 97301, United States. Tel.: +1 503 586 0127; fax: +1 503 586 0192.
E-mail address: forensictrauma@gmail.com (M.D. Freeman).

1752-928X/$ - see front matter © 2009 Elsevier Ltd and Faculty of Forensic and Legal Medicine. All rights reserved.
doi:10.1016/j.jflm.2009.09.006

New approaches to causal determinations in a forensic setting are carefully scrutinized by the opposing side, and occasionally rejected by courts as unsound. Recent court decisions have called into question the applicability of biomechanics in forensic settings, with one judge writing that an expert's opinion in a traffic crash injury case "*should be excluded for the reason that there is no scientific or medical basis for the use of biomechanics, engineering or accident reconstruction as a means to determine whether or not an injury has occurred as a result of a crash; this type of data is generally used to explain injury in an experimental setting rather than to explain away injuries that have already occurred*".[4]

It is reasonable to investigate the accuracy of such a methodology, since Forensic Biomechanics uses indirect measures such as applied force to serve as a proxy for injury risk (a population-based parameter), which in turn is used to assess the probability of injury presence (an individual characteristic). Epidemiologic observational study of injury is the gold standard for injury risk assessment[5] and measures intended to serve as a construct for risk that don't directly measure injury occurrence should be adopted and applied with appropriate caution. Conversely, a new methodology should not be rejected out of hand solely because it is new.

In the present paper, we describe a systematic assessment of the accuracy of Forensic Biomechanics, that is, how often such an analysis would be expected to be correct vs. incorrect. As part of our analysis, we examine the context of Forensic Biomechanics; whether it is used to determine "how" or "whether" an injury resulted from a traumatic event, or if manipulation of a hypothetical variable such as seat belt use would have made a difference in the injury outcome. Finally, we discuss how the epidemiology of injury plays an important but often hidden or neglected role in Forensic Biomechanics.

2. Injury and causation

Injury differs from disease largely by how rapidly symptoms are developed after an injurious event or exposure. For example, a fall down the stairs that results in a broken wrist is an injury, but carpal tunnel syndrome resulting from repetitive use of a key board is a disease.[6] In the former, the symptoms are temporally attributable to the fall as they occur immediately afterwards, whereas in the latter the symptoms may start in the middle of the night after 6 months of high intensity computer use. The determination of causation for a wrist fracture following a fall down the stairs doesn't require special expertise or knowledge, but the relationship between carpal tunnel syndrome and repetitive use of the hands is one that is not as easily recognized. For this reason, issues of injury causation are ordinarily decided by clinicians following generally accepted methods of causal determination based primarily upon the history of the event as related by the injured patient. This interaction helps establish the nature of the event and the temporal relationship between the event and the onset of symptoms. Clinical causation depends largely on the truthfulness of the injured party, and it is not unreasonable to question historical attributions of injury to an event from which compensation for injury can be recovered. Conversely, it cannot be said that just because compensation can be recovered for an injury that an injured party is likely to lie about the cause of his injuries. Forensic biomechanical injury assessment does not occur in a vacuum; all litigated cases have medical evidence of injury and the determination by a licensed clinician that the injury resulted from the injury mechanism. The following two *a priori* assumptions are generally valid in all cases in which such an analysis is requested:

(1) The attributed injury mechanism, a traffic crash for example, has occurred. It is assumed that the preliminary work of investigating the claim for an overtly fraudulent act, such as staging a collision, has already been performed, as insurance companies have Special Investigation Units (SIU) that are dedicated to this type of investigation and do their work long before litigation experts in Forensic Biomechanics are called in.[7]

(2) The injuries diagnosed by a claimant's physicians are real. This assumption is based on the work that would already be done by an insurer's SIU or others in investigating and identifying overtly fraudulent medical practices before a forensic biomechanical assessment is requested by a plaintiff or defendant. In some injury litigation claims a claimant may be found to be malingering (faking) or exaggerating an injury by an adverse medical examiner but the dispute as to whether the injury diagnosis is valid cannot be augmented by a biomechanical analysis. The assessment relates only to the causal relationship between the injury mechanism and the injury, and is unrelated to diagnostic validity issues. For example, if an injury was found to be physically impossible in a collision by a biomechanical analysis, say a traumatic amputation of a limb in a low speed collision, it is the injury *cause* rather than the injury *diagnosis* that would be disputed (an amputated limb would be difficult to fake). This is not to say that a biomechanical assessment of injury risk that resulted in a low injury probability would not fit with an explanation of malingered injury complaints, just that the determination of injury risk is independent of the injury or malingering diagnosis.

3. Forensic Biomechanics and probability

The two preceding assumptions dictate that determinations of probability based upon biomechanical assessments of injury risk must be approached with some caution when they are applied in a forensic setting. If it is assumed that the injury event occurred and the injury is real then it is reasonable to ask if a determination of biomechanical injury risk is even relevant to a determination of causation. As an example of how biomechanical injury risk can be misleading when making absolute determinations of injury causation, Tencer et al. described a cohort of 20 seat belt restrained crash victims with femur fractures that resulted from relatively low collision forces.[8] All of the crashes had been reconstructed for speed and a biomechanical analysis of the injury forces was performed following a vehicle interior inspection. The authors selected the cases because, when compared with experimental studies of femur fracture tolerance, most occurred at force levels equating to a biomechanical risk of fracture of 10% or less (this would correspond with a Factor of Risk of injury, as described by Hayes et al., of 0.1. In fact, two cases of fracture occurred at 1% risk and one case at 0% risk (Φ of 0.01 and 0.00, respectively). The low risk did not translate to low probability of injury, however, as all 20 subjects were injured, an unexpected but undisputed outcome. As a technique that is intended to discriminate between injury and no injury based on injury probability assessment, forensic biomechanical injury evaluation would lead to the erroneous conclusion that all of the fractures were improbable or even impossible. On the other hand such an evaluation could help explain why an injury that was deemed unlikely had occurred (e.g. Tencer et al. attributed the injuries to muscle forces acting along with the crash forces). The preceding example illustrates one of the pitfalls of an injury evaluation methodology that does not consider actual evidence or observation of injury in arriving at a conclusion of whether the injury was more probable than not.

This last point raises another issue to consider when evaluating the various applications of forensic biomechanical assessment of injury risk. All forensic testimony is given as "more probable or

likely than not" or as a "reasonable probability;" relatively interchangeable terms that serve as a quantifiable threshold that must be exceeded before the testimony is admissible.[9] Thus the expert must be "more than 50% certain" that the opinion is correct. Using probabilistic language for such testimony is somewhat of a mischaracterization of an internal process of the expert, who has opined that he is more certain than not that his opinion is accurate or true, regardless of the methods used to arrive at the opinion. When the opinion is scientific rather than medical, what constitutes a reasonable probability may result from an analysis of data, often from epidemiologic study. For example, if an alleged act of malpractice that altered the stage or projected aggressivity of a cancer was found to result in an increase in the probability of death within 5 years of more than 50% (from 5% to 10% for example), the conclusion that the plaintiff's life had been shortened could be given as a reasonable probability. In assigning a threshold value for injury probability vs. injury improbability, Hayes et al. create criteria for injury that can lead to erroneous conclusions. For example, if one interprets a reasonable probability of injury to be a Factor of Risk assessment that is 0.5 or greater, the application of this threshold to the cohort of 20 crash-related femur fracture cases described earlier would indicate that 16 of the 20 fractures did not occur, as only four were found to have a risk of more than 0.5. Some of this difficulty in translation comes from the conversion of a continuous variable (injury risk, 0–1 range) into a dichotomous variable (injury not present $\leqslant 0.5$, injury present >0.5), and some of it comes from potential limitations on the extrapolation of data that is used to assess biomechanical injury risk.

As an example of the limited utility of some experimental and observation data for application in forensic venues and the problems introduced by dichotomization of such data, Hayes et al. describe a study by Newman et al. of National Football League (NFL) players who sustained concussions (mild traumatic brain injury) during plays that were filmed from multiple angles allowing for a reconstruction of the forces of the injury event.[10] The logistic risk curve that was fit to the data is displayed in Fig. 1.[11] The authors use the data from this study to estimate a Factor of Risk of concussion for one of their traffic crash case studies.

The graph in Fig. 1 demonstrates calculated Head Impact Power for nine professional football players who were found to have sustained a concussion, in comparison with 15 players who were impacted and not injured. The 50% risk level equates to a HIP_m of 12.8, a value that divides most, but not all of the data into injured and not injured. It is worth noting that 2 of the 9 (22%) players' concussions occurred at values below the 50% threshold (HIP_m of 8.9 and 12.2). Likewise, one of the 15 uninjured players (7%) sustained a HIP_m of 19.7, equating to a greater than 0.90 concussion risk, with no apparent ill effects. As an illustration of difficulties with the dichotomization of continuous data, had the nine injured football players been diagnosed with a concussion following a fall while at work or in a traffic crash, a forensic biomechanical analysis of the injuries that used the 0.50 point on the risk curve as a cut off would have resulted in 1 in 5 having their claim of injury rejected as "improbable".

As another example that raises questions about dichotomization of continuous data, Hayes et al. refer to the Neck Injury Index (N_{ij}) several times as a basis for assessing the probability of injury of various severities to the cervical spine. The N_{ij} is a metric used for evaluating the performance of airbags relative to the loads they may produce in the cervical spine.[12] In Fig. 2, a graph depicts the risk of AIS 2 injuries (an Abbreviated Injury Scale) of the cervical spine by N_{ij} (a nondimensional metric). AIS 2 injuries are considered "moderate" in severity, and include disc injuries and fractures without spinal cord involvement.

The probability of an AIS 2 injury at an N_{ij} of 0.5 is 18.9%, meaning that the graph predicts that if 100 people are subjected to an N_{ij} of 0.5 approximately 19 will sustain an AIS 2 cervical spine injury. If the threshold of a 0.5 N_{ij} is used to determine the probability that a cervical disc injury resulted from a traffic crash in an individual the test will be wrong one time for every four times it is correct.

This error rate for the N_{ij} as a forensic test of injury causation given above assumes that the risk curve can be accurately applied to all occupants in all types of collision, an assumption that is not necessarily valid. The N_{ij} risk curve is derived from a risk curve originally derived from porcine neck testing.[13,14] The curve was then matched to injuries observed in real world non-rollover frontal crashes by extrapolating from the N_{ij} measured in male crash test dummies generated during frontal barrier crash testing conducted at 35 mph. All of the crash tests involved airbag deployments and belted dummies. Although scaling and extrapolation can be performed to some extent, because the N_{ij} has not been validated for injury mechanisms outside of frontal traffic collisions any use of the metric for other injury mechanisms should be approached with caution.

There is certainly a justification for the development and application of new methods of evaluating the truth of assertions or allegations in disputed matters. A plaintiff in a civil litigation may have a completely legitimate claim of injury, but it cannot be denied that there exists a financial incentive to exaggerate or fabricate injuries and that exaggeration and fabrication occurs in some cases. Likewise, a defendant may have a legitimate reason for denying compensation to a claimant, but it must be acknowledged that there is a financial benefit to a defendant who prevents a legitimately injured plaintiff from being awarded compensation, and that in some cases claimants with real injuries are deprived of compensation. The fact finders' dilemma is how much weight they should give to the testimony of plaintiffs regarding their injuries, or to their treating physicians who may only be relying upon the plaintiff's history for a causal determination. A test that is intended as an impartial determinant of the truth in a disputed matter is an

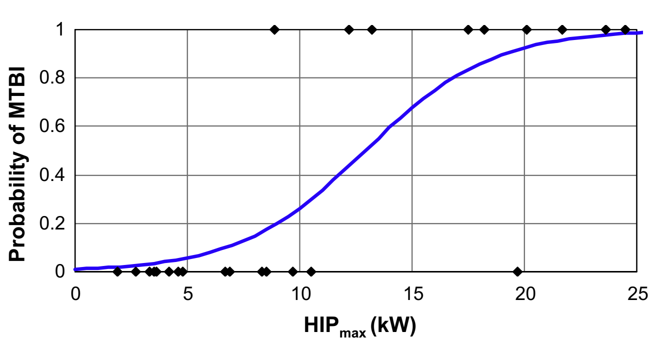

Fig. 1. Probability of concussion based on incidences of Mild Traumatic Brain Injury (MTBI) associated with the calculation for maximum Head Impact Power (HIPmax) (adapted from Newman et al.[11]).

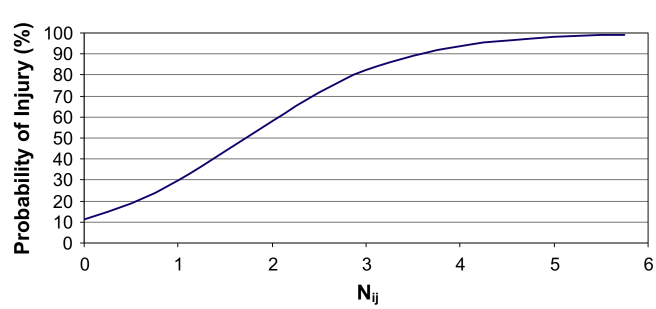

Fig. 2. Probability of an AIS 2 injury as predicted by the Neck Injury Index (N_{ij}) (adapted from Eppinger et al.[12]).

ideal solution for the fact finder only if the test can be shown to be accurate.

3.1. Assessing the error rate of forensic tests

In 1993, the United States Supreme Court issued an opinion in a case called *Daubert v. Merrell Dow Pharmaceuticals Inc.* The opinion set new standards for evidentiary hearings in the United States, in which the judge acts as a gatekeeper for proposed scientific testimony.[15] The Daubert decision laid out certain criteria for the admissibility of testimony that was based on new or novel methods, including four general standards considered by the Court to represent the basis for a valid scientific method, including:

(1) Whether the technique had undergone empirical testing; the theory or technique must be shown to be falsifiable, refutable, and testable;
(2) that the technique had been subjected to peer review and publication;
(3) that there is a known or potential error rate of the method; and
(4) whether or not the method is generally accepted by a relevant scientific community.

The Daubert standard has been used to challenge the admissibility of forensic biomechanical testimony on a number of occasions, and such testimony has been precluded as lacking in a scientifically sound basis on dozens of occasions in recent lower[16–22] and appellate court decisions[23–26], often because the expert cannot quantify the degree of error inherent in the analysis.

The error rate of a method, the 3rd listed criteria for an acceptable methodology under Daubert, is a good means of quantifying the accuracy of a method that serves as a test for truth in a disputed matter. When evaluating test accuracy, it is helpful to first understand how a test may be inaccurate. If a test is like a rifle with a scope mounted, then an accurate test would yield results like the target in Fig. 3a, in which the shooter knows that when the bulls eye is in the crosshairs of the scope that the bullet will strike the bulls eye. If there is too much random error in the test, in the rifle analogy this might be a scope that is mounted too loosely to the rifle, then the results will be unpredictable, like in Fig. 3b. If there is systematic bias in a test, if the scope is mounted to the rifle at an angle for example, then the results will be predictably wrong in a certain direction, as depicted in Fig. 3c.

3.2. Conditional probabilities and Bayes' Law

The purpose of any forensic test is to "condition" the probability of a particular outcome or result. For example, if the issue of interest is the probability of a particular injury "A" following a traffic crash, depicted symbolically as $P(A)$, then a conditioned probability would be the probability of injury given the presence of another factor; a positive test "B" for example. This conditional probability is depicted symbolically as $P(A|B)$; the probability of injury A given the positive test result B. An error that may occur when evaluating a conditional probability is that the assumption is made that the terms are reversible; that $P(A|B) = P(B|A)$. This error is called a Conditional Probability Fallacy.[27] An example of the Conditional Probability Fallacy is the conclusion that "90% of Spaniards speak Spanish" is equally true as "90% of Spanish speakers are Spaniards". As applied to a forensic test, the Conditional Probability Fallacy occurs when it is erroneously concluded that the probability that a test will be positive when a condition is present is the same as the probability that a positive test means the condition is present (symbolically represented as $P(\text{test positive}|\text{condition}) = P(\text{condition}|\text{test positive})$. As an example, one could devise a test for guilt that was based on body temperature; if the body temperature was above 50 °F the test would be positive for guilt. The test would be positive 100% of the time that the defendant was guilty, but quite obviously the defendant would not be guilty 100% of the time the test was positive. This type of error is also known as the Prosecutors Fallacy.[28]

The way to avoid a Conditional Probability Fallacy is through the application of Bayes' Law, the principles of which are critical to the evaluation of the potential error rate of a forensic test. Most simply stated, Bayes' Law allows for a more precise quantification of the uncertainty in a given probability. As applied in a forensic setting, Bayes' Law is a method of finding out what we want to know given what we know.[29] Bayes' Law is named for the essay by Reverend Thomas Bayes (1702–1761) on the statistical analysis of probability[30], and over the past 250 years subsequent authors have further defined and refined Bayes' original propositions. In more recent years Bayes' Law has been used in forensic settings primarily to quantify probabilities associated with DNA testing, however its use has expanded to virtually any disputed issue that requires the quantification of probability and its complement, uncertainty (uncertainty = [1 − probability]).[31,32]

At its most simple, Bayes' Law is stated $P(A|B)$ in which the probability of A is dependent upon condition B, vs. the probability of A absent any conditions. For example, the probability of drawing an ace of hearts from a deck of cards is 1:52; however the probability of drawing an ace of hearts given that only red cards can be selected is 1:26, the probability of drawing the ace of hearts when only hearts are drawn is 1:13, and the probability of the ace of hearts when only aces are drawn is 1:4.

When used to evaluate the accuracy of a test the application of Bayes' Law allows for the consideration of the most important conditions that can influence the ability of the test to arrive at a correct answer. As an example, we can hypothesize a test for the presence of illegal drugs that has a sensitivity rate of 0.90 (it correctly identifies test subjects with drugs in their system as positive 90% of the time) and a specificity rate of 0.80 (it correctly identifies test subjects with no drugs as negative 80% of the time). The sensitivity rate is the same as the *true positive* rate (0.90), and the complement

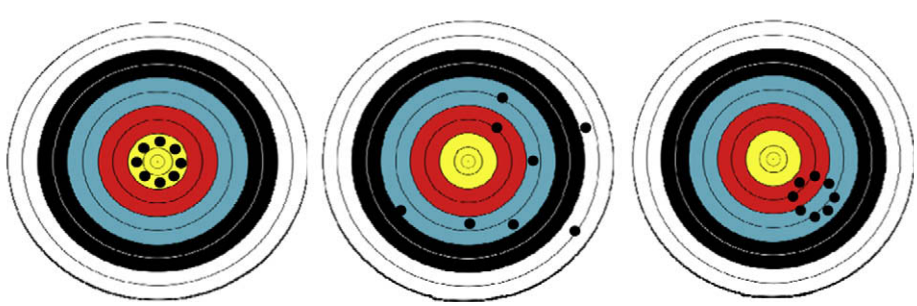

Fig. 3. 'Accurate' test results (a, left bullseye); 'Inaccurate' test results due to random error (b, middle bullseye); 'Inaccurate' test results due to bias (c, right bullseye).

of the specificity rate (1 − specificity) is the *false positive* rate (1 − 0.80 = 0.20). The ratio of the true positive rate to the false positive rate is also known as the Likelihood ratio for the test:

$$\text{Likelihood ratio} = \frac{\text{true positive rate}}{\text{false positive rate}}$$

3.3. Error Odds

In this paper, we introduce a simple application of Bayes' Law for assessing the degree of uncertainty in a positive test result, called the Error Odds (notated as O_E). In Bayesian terminology the Error Odds assessment is also known as the post-test or posterior odds. The result of the Error Odds assessment is the ratio of true positive to false positive tests given the expected "base rate" or frequency of the condition of interest in the population like the test subject (also known as prevalence). The equation for the Error Odds is as follows:

$$O_E = \frac{\text{true positive rate}}{\text{false positive rate}} \times \frac{\text{pre-test probability}}{(1 - \text{pre-test probability})}$$

The Error Odds assessment is unique in that it is designed for application in a forensic setting, and as such is only applicable to positive test results (true and false negative rates are not considered with the Error Odds). The rationale for a test validity measure that only considers true or false positives arises from the fact that the results of all forensic tests are ultimately presented as positive outcomes that support one or the other side of a disputed issue, regardless of the test outcome. As an illustration, a blood alcohol test administered in a drunk driving investigation that indicated the presence of alcohol would be presented by the prosecution to a jury as a positive test for guilt. The same test, if negative for alcohol presence, would be presented to a jury as a positive test for innocence. Thus the Error Odds assessment of a positive test result can be briefly described as the ratio of the rate of correct positive results to the rate of incorrect positive results. The Error Odds is intended as a "snapshot" of the degree of uncertainty in a test result. Note that the results of Error Odds calculation is actually the odds *against* error; this reversal is necessary so that a threshold for an acceptable level of error is a whole number below which the odds of error can be considered to be unacceptably high for application in a forensic venue.

As an illustration of the application of the Error Odds the results of drug testing of two hypothetical populations can be assessed for comparative validity. For the example, the true and false positive rates of the previously described drug test can be used (0.90 and 0.20, respectively). In order to complete the assessment a final piece of information is needed, and this is the prevalence or pre-test probability of drugs in the tested populations. The hypothetical test populations are: (1) a group of 100 prisoners recently incarcerated for a drug related crime in which drugs have been historically found in 95 of the subjects (0.95 drug prevalence) and (2) a group of 100 6th grade students in which drugs have been previously found in only 5 of the subjects (0.05 drug prevalence). Using the 0.9 true positive and 0.2 false positive rates the Error Odds assessment for the two populations is presented diagrammatically and in equation form in Figs. 4 and 5.

The calculation yields Error Odds of 86 and 0.24, or 86 correct test results for every 1 incorrect result for the prisoners and 0.24 correct results for every 1 incorrect result for the 6th graders. The Error Odds would be the same for any single individual test result. Note how sensitive the results of the accuracy test are to the prevalence of the condition of interest; even though drug presence is only 19 times more prevalent among prisoners than among 6th

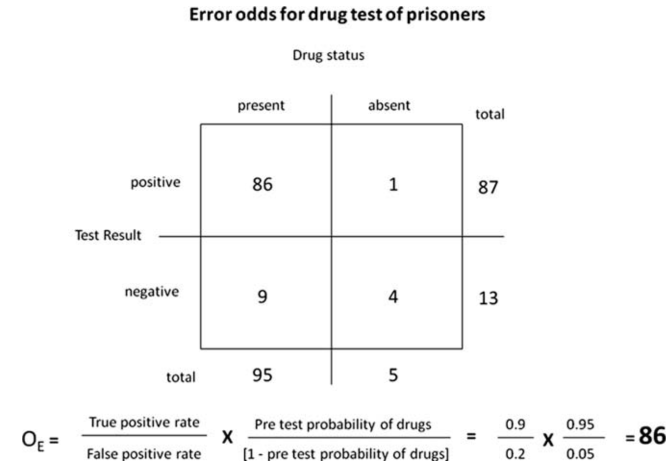

Fig. 4. A 2 × 2 contingency table summarizing the example analysis of the Error Odds associated with drug use in a hypothetical prison population.

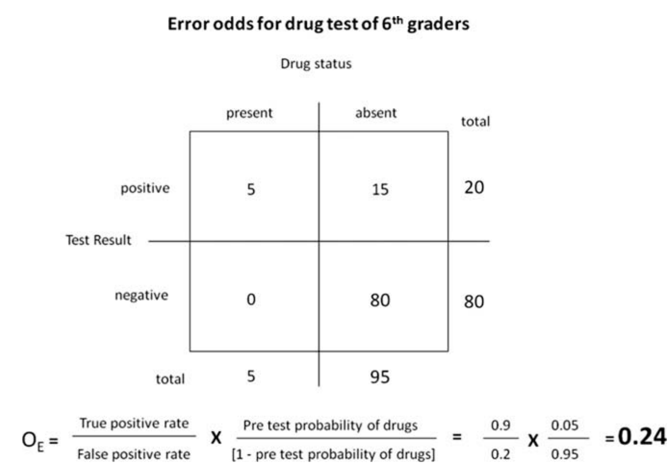

Fig. 5. A companion 2 × 2 contingency table supporting the example analysis of the Error Odds associated with drug use in a hypothetical elementary school population.

graders (0.95/0.05), the test is correct 358 times more often among prisoners than among 6th graders (86/0.24).

Also note in the example how the application of Bayes' Law avoided a Conditional Probability Fallacy or Prosecutor's Fallacy. The probability that the test would be positive given the presence of drugs was 0.9 for both groups however the conditioned probability that drugs would be present given a positive test was widely disparate between the two populations.

It is difficult to set an Error Odds value that qualifies a test result as "acceptable" for forensic testimony. It has been suggested that when Likelihood ratios are used to evaluate medical diagnostic test results a value of 10 is a minimum threshold from which to conclude that a patient has the condition for which he was tested.[33] In a similar vein, a lower threshold of 10 correct tests for every one incorrect test has been previously proposed for the evaluation of evidence in criminal matters.[34] With reference to a threshold ratio of 10, the matrix presented in Table 1 illustrates the Error Odds for a hypothetical forensic test with a true positive rate of 100%, given prevalence and false positive rates ranging from 0.05 to 0.95. The bolded cells on the right side of Table 1 are those that exceed 10. Note that if the pre-test probability of the condition is less than 0.5 the Error Odds do not exceed 10, regardless of the false positive rate.

Table 1
The breadth of Error Odds values assuming a true positive rate of 100%.

False positive rate	Pre-test probability						
	0.05	0.1	0.25	0.5	0.75	0.9	0.95
0.05	1.1	2.2	6.7	**20**	**60**	**180**	**380**
0.1	0.5	1.1	3.3	**10**	**30**	**90**	**190**
0.25	0.2	0.4	1.3	4	**12**	**36**	**76**
0.5	0.1	0.2	0.7	2	6	**18**	**38**
0.75	0.07	0.1	0.4	1.3	4	**12**	**25**
0.9	0.06	0.1	0.4	1.1	3	**10**	**21**
0.95	0.06	0.1	0.4	1.1	3	9	**20**

3.4. Error Odds modifiers

The presence of either bias and/or random error may increase the probability that a forensic test will yield an erroneous result by limiting the extrapolability of the data to the circumstances of the disputed matter, but the magnitude of effect may be difficult to estimate. Forensic Biomechanics relies in large part on experimental studies of animal, cadaver, and occasionally human volunteer subjects. Real world human tolerance levels demonstrate far greater variance than what can be produced experimentally for the simple reason that experiments cannot reproduce the breadth of variability inherent in the population, nor can they reproduce the variability of circumstances in which injuries occur. Fig. 6 illustrates a series of hypothetical Gaussian distributions that demonstrate the variability and quantity of experimentally derived data vs. the extent of variability in the general population. Error results when the experimental data is used as an index of risk for what may be observed in the general population.

The publication that forms the basis for the human tolerance metrics used in the Federal Motor Vehicle Safety Standards for automotive design and testing by the US Department of Transportation characterizes the extrapolability of human tolerance data to real world settings as follows[35]:

> "Such [tolerance] specifications are beyond the state-of-the-art in biomechanics except perhaps for a few academic situations. There are several difficulties which prevent a ready establishment of human tolerance levels. First, there are differences in judgment as to the specific degree of injury severity that should serve as the tolerance level. Second, large differences exist in the tolerances of difference individuals. It is not unusual for bone fracture tests on a sample of adult cadavers to show a three-to-one load variation. Presumably, variations of at least this magnitude exist in the living population. Finally, most tolerance levels are sensitive to modest changes in the direction, shape, and stiffness of the loading source. The above considerations indicate that complete and precise definitions of human tolerance levels will require large amounts of data based on controlled statistical samples. Only in this way can the influence of age, size, sex, and weight be comprehensively assessed and only in this way can mean loads and statistical measures of scatter be linked to specific tolerance levels."

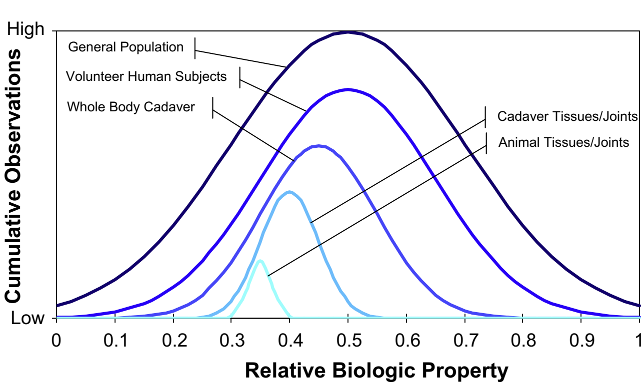

Fig. 6. A set of hypothetical Gaussian distributions demonstrating the variance, mean, and source of experimentally derived sample data compared with the general population.

The caution advised in the preceding statement refers to the effect of both bias and scatter on the extrapolation of tolerance specifications to real world occurrences as boundary conditions. We do not interpret this to mean that extrapolation of tolerance data cannot be accomplished in a forensic setting; rather, limitations and error should be identified prior to determining whether such data are suitable as a basis for a forensic opinion. This means that the influence of bias and random error should be evaluated in the methods and data used to arrive at the opinion.

3.5. Bias

When extrapolating from a study population to an individual in a Forensic setting it is ideal that the reference population be identical to the individual in every possible way in order to minimize bias. As this situation is a rarity, the next best approach is to characterize and quantify how the study population differs from the individual in aid of a determination as to whether the extrapolation is feasible. As an example, the study mentioned earlier by Newman et al. describing the injury mechanisms by which NFL football players sustained concussions raises the obvious question of whether the study population is more resistant to injury than the average motor vehicle occupant. The result is that values at which injury occurred in the study population are inflated, relative to the average motor vehicle occupant. The difficulty arises in assessing this potential difference; it would be a relatively simple task to quantify non-injurious proxies of injury resistance such as muscle strength in NFL players vs. the average motor vehicle occupant, however the degree of force required to cause a concussion is a different matter that may or may not be related to physical hardiness. It is reasonable to conclude however, that the extrapolation of data from a study group that consists of the most elite and physically hardy members of the population to an ordinary individual cannot occur without some degree of scaling, and that the failure to do so increases the probability of erroneous conclusions.

3.6. Random error

Lack of sufficient study numbers is the primary cause of random error, which only becomes an issue when the data are used as an exclusive rather than inclusive boundary condition. Referring back to the Newman et al. study on head injury, for example, the authors observed concussions at HIP_m levels as low as 8.9 kW. A reasonable conclusion from these data is that a concussion observed in any individual at a HIP_m of 8.9 kW or greater is consistent with the study data. The converse of this conclusion, that a concussion observed at a HIP_m of *less* than 8.9 kW is *not* consistent with the study data, would be erroneous to the extent that the limited number of injured subjects (here nine players) is insufficient to draw a sharply delineated boundary below which it can be said that injury cannot occur. The following analogy helps explain how the type of opinion (inclusive vs. exclusive) dictates the sensitivity of a test result to study numbers. If one had a box full of 100 balls that contained an unknown quantity of either red and/or black balls, and the balls could only be examined one at a time, if a single red ball was drawn one could conclude that red balls are present in the box at some unknown frequency. This would be a reasonable *inclusive* conclusion based on the limited sample. It

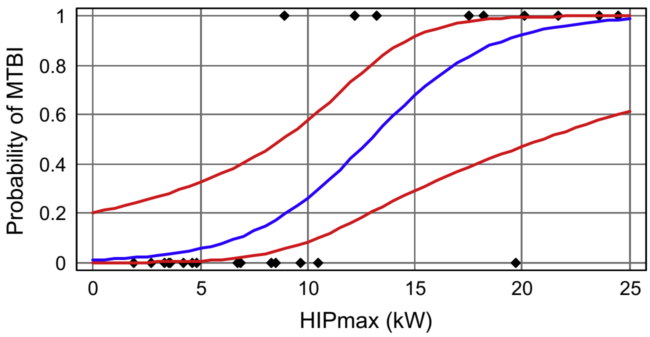

Fig. 7. Similar to Fig. 1, a 95% Confidence Interval (the red lines above and below the middle blue line) have been added to demonstrate the variability of MTBI associated with HIPmax. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

would be an error to conclude that black balls are impossible, however. This would be an *exclusive* conclusion that is only supported by the data if one ignores the effect of random error, which dictates that the sampled red ball may have been the only one in the box, and that rest of the balls in the box may be black. Only by increasing the sample size can one decrease the effect of random error.

One way to account for the effect of study numbers on random error (but not bias) is to present measures of central tendency such as a mean and regression with confidence intervals that take into account the variation among the data. As applied to the previously presented Newman et al. data[11] the 95% confidence interval for a 50% risk of concussion or mild traumatic brain injury (MTBI) among NFL football players would range from approximately 9 kW to 21 kW, as demonstrated in Fig. 7.

4. Case studies in Forensic Biomechanics; an Error Odds analysis

Hayes et al. describe their application of Forensic Biomechanics to three cases in which injury resulted from a traffic crash and four cases in which injury resulted from a fall. While it is estimated by the US Centers for Disease Control that there are more than twice as many fall-related injuries evaluated and treated in hospital emergency departments than traffic crash-caused injuries,[36] a much smaller proportion of fall injuries are litigated relative to traffic crash injuries. Taking into account the frequency of injury vs. the frequency of litigation the rate of crash injury litigation is estimated to be more than eight times greater than for falls,[37] and thus it would be expected that forensic biomechanical analysis of injury risk would be encountered significantly more frequently in traffic crash litigation than in slip and/or fall litigation.

4.1. Estimating Error Odds input values for various types of forensic biomechanical opinions

In the following section we present an Error Odds assessment of the seven case studies described by Hayes et al. as a means of demonstrating the utility of the methodology. This section is organized by the type of forensic biomechanical conclusions associated with each case. The cases were separated into three categories, relative to how the medical evidence of injury fit with the reconstructed circumstances of the injury mechanism. There are four cases (3 crashes and 1 fall) in which the major conclusions were deemed to be *explanatory*, in that they provided an explanation for how an injury occurred when the injury mechanism was not apparent (such as with an unobserved fatal fall). There are two cases (1 crash and 1 fall) that included an *alternate hypothetical* scenario in which the theoretical effect of an injury reduction mechanism such as a seat belt was evaluated and compared to the actual event. The third category of conclusions, *refutative,* is illustrated with a traffic crash case in which a diagnosed injury was deemed unlikely to have resulted from the collision.

This section is organized as follows; first, a synopsis of the case details from the Hayes et al. paper is presented. This is followed by estimates of the three elements needed for the Error Odds assessment; the pre-test probability and the true and false positive rates of the test. Values are reported in ranges when appropriate, and estimates have been adjusted to favor a lower error rate in order to maximize the Error Odds value. Estimates of the effect of bias and error on the values are also presented when warranted. The results of the Error Odds calculations for all of the cases are presented in Table 2.

The most difficult value to estimate for the Error Odds analysis was the pre-test probability of the condition of interest. This value is not the same as simply determining the prevalence of the injury given the circumstances, however. A Bayesian (conditioned) perspective is helpful to understand the pre-test probability necessary for an Error Odds assessment. As an illustration, if 100 occupants are exposed to a 10 mph collision, there may be only 20 who are injured, resulting in an injury prevalence of 0.20. If all of the 20 injured occupants made a claim for injury then the prevalence of injury among the claimants would be 1. If 10 of the 80 uninjured occupants made a false claim for injury in addition to the 20 injured occupants then the frequency of real injury among all claimants would be 20 out of 30, resulting in a pre-test probability of 0.67. It is this value that would be used for an Error Odds assessment, rather than the 0.20 or 1, both of which would result in a Conditional Probability Fallacy.

Thus, in the proceeding analysis section, the pre-test probabilities for some of the Explanatory and Refutative cases were arrived at not by asking "*how often do these injuries occur*" but rather "*how*

Table 2
Input values supporting Error Odds calculations for the re-examined case studies.

Analysis type	Case no.	Description	Pre-test probability (PTP)	True positive (TP) rate	False positive (FP) rate	Likelihood ratio (TP/FP)	Pre-test odds (PTP/[1 − PTP])	Error Odds (O_E)
Explanatory	1	Traffic crash neck injury	0.93	0.5 (Any injury) 0.5 (long term)	0.5 (Any injury) 0.5 (long term)	0.5 (Any injury) 0.5 (long term)	13.3	**18.5 (Any injury) 30.2 (long term)**
	2	Slip and fall in ramp	0.5	0.89	0.64	1.39	1	**1.4**
	3	Fall from balcony	0.35	0.35	0.65	0.54	0.54	**0.3**
	4	Fall on trampoline	0.53	0.53	0.38	1.39	1.13	**1.6**
Hypothetical	*5*	*Traffic crash belted vs. unbelted*	*0.8 (Head) 0.53 (neck)*	*0.8 (Head) 0.53 (neck)*	*0.8 (Head) 0.53 (neck)*	*0.8 (Head) 0.53 (neck)*	*0.8 (Head) 0.53 (neck)*	**16 (Head) 2.4 (neck)**
	6	*Exercise bike fall no safety vs. safety*	*0.99 (No safety) 0.5 (safety)*	*0.99 (No safety) 0.5 (safety)*	*0.99 (No safety) 0.5 (safety)*	*0.99 (No safety) 0.5 (safety)*	*0.99 (No safety) 0.5 (safety)*	**9800 (no safety) 1 (safety)**
Refutative	7	Traffic crash head injury	0.33	0.65	0.35	1.86	0.49	**1.4**

often are observed injuries real/not real?" The estimated pre-test probability is conditioned by the desired positive test result. This concept is explained in greater detail in the case analyses below:

4.2. Explanatory cases

(1) A traffic crash injury is described in which a minor severity (AIS 1) neck injury was deemed consistent with the injury potential of the collision.
- Pre-test probability = 0.93. This value was estimated to be the proportion of litigants who have been diagnosed with a minor neck injury who have a real injury. The value was derived from a literature based assessment of rate of malingering and fraud (1 − [fraud and malingering rate] = rate of real injury). The highest estimate of malingering or fraud reported in the literature was 33% and the lowest was 7.5%[38–43] and thus the pre-test prevalence of real injury was estimated to range from 67–92.5%. The lowest rate of malingering was selected for the Error Odds assessment to favor the highest Error Odds for the test.
- True positive rate = 0.5 (any injury), 0.5 (long term injury). The authors' reconstruction of the collision indicated a rearward speed change of 10.7 mph, and a Neck Injury Criterion (NIC) score of 25 m^2/s^2 was inferred from loads observed in the computer model. The injury potential of the collision was assigned a Factor of Risk of injury of 1.0 and 1.7 for long term and short term injury, respectively, based upon the NIC values reported by Boström et al. for 50 occupants with AIS 1 neck injuries from real world collisions (average 15 m^2/s^2), of which 11 lasted more than 6 months (average 25 m^2/s^2).[44] The injury rate at and above the neck injury threshold of 15 m^2/s^2 was 50% (25/50 injured above the threshold). The injury rate at and above the long term injury threshold was also 50% (6/11 injured). The small numbers in the sample increased the random error of the estimate, but this was not an issue as the conclusion from the data was *inclusive* rather than *exclusive* (see the previous discussion under Random Error).
- False positive rate = 0.36 (any injury), 0.22 (long term injury). In the reference population described by Boström et al., at the injury threshold of 15 m^2/s^2 there were an estimated 44 out of 122 (36%) uninjured occupants and above the 25 m^2/s^2 threshold there were an estimated 27 of 122 (22%) uninjured occupants.

(2) A fall injury is described, which took place on a declining ramp. The claimant was a 61 year old female in whom Hayes et al. determined that a diagnosed fractured patella was more likely to have resulted from a forward-falling trip rather than a slip as claimed by the claimant, as the latter scenario would have been more likely to have produced a backwards or sideways fall.
- Pre-test probability = 0.50. This value was estimated to be the proportion of litigants who report an inaccurate history of a fall either intentionally or mistakenly. There are no reference values from which to base this estimate, and thus the approximate range was derived from the values given previously for malingering and fraud (7.5–33%) plus an additional 50% to allow for error (12–50%). The highest rate was selected to favor a higher Error Odds calculation.
- True positive rate = 0.89. As the basis for their conclusion that the fall was more probably the result of a trip rather than a slip Hayes et al. referred to study by Smeesters et al. of experimentally produced slips and trips that produced a 93–100% rate of forward falls associated with trips, compared with a 72–79% rate of sideways or backwards falls resulting from slips.[45] Absent the effects of bias this would equate to a true positive rate of 0.93–1, and a false positive rate of 0.21–0.29 (the complement of the rate of sideways or backwards falling from a slip was the rate of forward falling). An evaluation of whether the results of the study were extrapolable, without scaling, to the circumstances and individual at the center of the dispute demonstrates a cause for concern. Smeesters et al. evaluated the fall patterns of 14 athletic and fit men and women (seven of each), average ages 22.2 and 20.7, respectively, by using a level platform equipped with mechanisms to induce slips or falls. The study participants, who were equipped with knee pads, were instructed to not attempt any recovery during the fall on the platform, which was padded in all directions the fall could occur. The results of the study for slips and trips, as presented in Table 1 on p. 312 of the paper, indicate a forward fall rate of 93% for trips at slow gait speed and 100% at fast gait speed. In the same table, the forward fall rate for slips at slow gait speed is reported at 21% and 64% for fast gait speed (it is not apparent from the paper where the 21–29% used by Hayes et al. came from). The small subject numbers and dissimilarity of the conditions and participants in the study relative to the fall circumstances and characteristics of the litigant indicate the probable influence of both random error and bias in the extrapolation of the study results to the disputed matter as a an exclusionary boundary. To take this influence into account in the Error Odds calculation we decreased the true positive rate by a factor of (1.5 × [1 – true positive rate]) so that the adjusted rate was 0.86.
- False positive rate = 0.64. The highest rate of false positives from the Smeesters et al. paper (the rate of slips that could produce a forward fall) was 0.64. To maintain the practice of underestimating, rather than overestimating the degree of error in the forensic biomechanical analysis this value was not adjusted, although it is as prone to random error and bias as is the true positive estimate.

(3) A fatal fall from a balcony is described. The decedent was a 57 year old man with a history of a seizure disorder in which it was determined that the cause of the fall over the balcony was the result of a seizure. Hayes et al. describe a computer model reconstruction of the kinematics of the fall, matched to the injuries sustained by the decedent. They describe their case question as "*whether or not the man could have fallen from the balcony as the result of a seizure*" (p. 76) but state their conclusions as "*the man fell from the balcony as the result of a seizure* (p. 77)". The shift from what was a *possible* or inclusive explanation for the cause of the fall to a *certain* or exclusive explanation is important from a forensic perspective, as the former is not admissible as evidence in most courts of law whereas the latter is.
- Pre-test probability = 0.35. There is nothing in the biomechanical analysis of the fall that would indicate that the initiation of the fall was a seizure vs. any other cause. The fact that the decedent had a history of seizures only makes the explanation of a seizure as a cause of the fall possible, however the success of the medical management of the seizure disorder, the recent past history of seizure frequency, and the type of seizures that the decedent suffered from are all more valuable determinants that would condition the probability that the seizure was the cause of the fall.[46] Research concerning the epidemiology of fall-related injuries in patients with seizure disorders indicates that,

although they are prone to suffer fracture more than three times as often as the population without a seizure disorder, only 35% of the excess fractures in seizure patients are associated with a seizure event.[47]

- True positive rate = 0.35. As the biomechanical analysis is not a test for seizure disorder the true positive rate is the same as the pre-test probability.
- False positive rate = 0.65. This is the complement of the true positive rate.

(4) A fall injury on a trampoline is described. The fall resulted in a cervical spine fracture and spinal cord injury in a 40 year old man in whom it was determined that the forces generated at the neck if the feet slipped out from under the body while gently jumping on the trampoline were the most probable cause of the injuries, as opposed to the theory that he was attempting to perform a back flip. As with the prior case of the fall from the balcony, Hayes et al. initially state their goal as inclusive; "…*whether a slip and backward fall-was a plausible mechanism for this injury* (p. 79)," but state their conclusions as exclusive; "…*the incident description provided by the subject was the likely mechanism for his injury and that he had not, as had been alleged, been attempting a back flip the first time he had set foot on a trampoline* (p. 80)."

- Pre-test probability = 0.53. While the forensic biomechanical analysis was used to establish that the force generated at the neck during a slip and fall were sufficient to generate supra-threshold AIS 3 and 5 forces in the neck using the N_{ij} compression values, this analysis could not be used to conclude that this injury mechanism was more probable than another mechanism (an attempted back flip) that could generate as much or more axial loading in the neck. As was the case with the previous seizure-fall case analysis, causal variables not addressed by the biomechanical analysis must be evaluated in order to properly condition the pre-test probability. The past behavior of the litigant (risky vs. conservative) is one consideration, and another is the epidemiology of trampoline-related injuries. In a study of 556 patients with trampoline-related injuries there were only eight cervical spine injuries, and only two of those were fractures.[48] For the entire study cohort 53% of the injuries were associated with an "awkward landing" on the trampoline, which would include a possible slip and fall mechanism.
- True positive rate = 0.53. As the forensic biomechanical analysis cannot additionally condition the probability of a slip and fall mechanism the true positive rate is the same as the pre-test probability.
- False positive rate = 0.38. Three of the eight cervical injuries occurred during an attempted somersault (the alternate scenario), a frequency of 37.5%.

4.3. Hypothetical cases

(5) A traffic crash fatality resulting from a high speed (74 mph) side impact collision and subsequent head injury is described. The risk of the observed critical severity (AIS 5+) head injuries is assessed for a belted and unbelted occupant. Hayes et al. conclude that the use of a seat belt would have likely eliminated the head injury but would have been likely to produce a critical neck injury instead.

- Pre-test probability = 0.8 (head injury), 0.53 (neck injury). Hypothetical scenarios require the estimation of both of the pre-test probabilities that are the source of the comparison. Hayes et al. do not provide a pre-test probability for the absence of a critical head injury in a belted occupant, leaving no means of assessing the accuracy of the conclusion that such an injury would have been *unlikely* had a belt been worn (in other words, given the magnitude of the collision, the injuries were likely regardless of restraint use). Thus for the Error Odds estimate the range of probability that a critical head injury would *not* have occurred in the restrained scenario was estimated to be low based on the high speed of the impact and the estimated HIC of 13,000 for the unbelted scenario, a value more than four times greater than the 99% risk threshold for critical head injuries as presented in Fig. 3 in the Hayes et al. paper. The result of this estimate was 0.5–0.8, a wide range meant to take into account the large degree of uncertainty in the estimate. The highest value was used for the Error Odds calculation. With regard to the neck injury risk, Hayes et al. reported the risk of critical neck injury as 53% for the calculated N_{ij} of approximately 3.0, and this was used for the pre-test probability for neck injury.
- True positive rate = 0.8 (head injury), 0.53 (neck injury). Although there is some potential for bias in the use of the N_{ij} for a side impact collision (it is derived from and validated with data on frontal collisions) the relatively large amount of data that it represents reduces the risk of random error and thus it was not adjusted. The basis for the true positive rate for the lack of critical head injury is stated above. Note that for this hypothetical scenario the pre-test probability is the same as the true positive rate.
- False positive rate = 0.2 (head injury), 0.47 (neck injury). These values are the complements of the true positive rates.

(6) The risk of a fatal head injury (an expanding subdural hematoma) is described for a 59 year old man who fell from an exercise bicycle onto a vinyl floor. The risk of the serious (AIS 3) severity injury was compared to the risk of the same severity of injury had the floor been padded, or if the man had been wearing a helmet or been seated on a recumbent (lower) bicycle. In their description of the case, Hayes et al. disclose that the decedent had been taking the anticoagulant medication Coumadin (generic name warfarin) prior to the fall. This case was the only one in which the authors did not reconstruct the forces of the fall. Instead they relied upon a fracture tolerance study of cadaver skulls in which it was determined that a HIC of 1000 was the threshold for fracture in the experimental population.[49] Hayes et al. noted that a subdural hematoma has an AIS severity level of three (serious) and that a HIC of 1000 corresponds to an AIS 3 or greater injury risk of 50–60%. The authors thus conclude that the HIC for the fall was at least 1000 because the presence of the subdural hematoma indicated that the HIC exceeded the risk threshold required to initiate a subdural hematoma (presumably 50%), but the HIC was not greater than 1000 because no skull fracture occurred. The authors did not provide any HIC estimates for the alternate scenarios, concluding only that floor padding would have prolonged head contact duration, a helmet would have attenuated impact force, and a recumbent bicycle would have reduced the fall height and resulting head impact force.

- Pre-test probability = 0.99 (no safety measures), 0.5 (safety measures). These values are the probability that the decedent would sustain an AIS 3+ injury without any safety measures in place, and the probability that

he would not have sustained an AIS 3+ injury if one of the safety measures been in place. The greatest complication to the analysis of the force required to cause the subdural hematoma seen in the decedent was his use of anticoagulant therapy, as this medication use is strongly associated with an increased risk of intracranial bleeding, and has been estimated to increase the risk of such bleeding by 7–10-fold.[50,51] Indeed, anticoagulated patients have been found to have sustained subdural hemorrhages following exposure to the relatively mild and typically non-injurious head accelerations that occur on roller coasters.[52] This leads to two difficulties with estimating a pre-test probability of injury for the forensic biomechanical analysis; first, the rationale employed by Hayes et al. in arriving at a HIC of 1000 (that a 50% risk of an AIS 3+ injury was the threshold for injury the decedent) cannot be extrapolated to an anticoagulated individual. As there is no lower boundary for the force required to cause the injury, then all that is left is the meaningless upper boundary of a HIC of 1000, below which a skull fracture would not be expected. What is known is that the fall and head impact did result in a subdural hematoma in the decedent, and thus whether the associated HIC was 1000 or 100, the pre-test probability for injury to this individual in this event is assigned a near certain probability of 0.99. The reduction of head impact force from an unknown amount would tend to reduce the risk of injury; however there is no way to know how much this would be, as the risk of injury may have been drastically reduced or not reduced at all by the safety measures. In order to complete the Error Odds analysis of this case study, however, the pre-test probability of head injury in the hypothetical scenario with safety measures is estimated to be half of the risk when no safety measures are present, or 0.50.

- True positive rate = 0.99 (no safety measures), 0.5 (safety measures). These rates are the same as the pre-test probabilities as the forensic biomechanical analysis could not additionally condition the probability of injury.
- False positive rate = 0.01 (no safety measures), 0.5 (safety measures). These values are the complements of the true positive rates.

4.4. Refutative case

(7) The authors describe a traffic crash case in which an unrestrained driver of a semi tractor-trailer was diagnosed with a concussion following a left frontal collision with another semi tractor-trailer. Hayes et al. conclude that the forces of the collision were insufficient to cause the concussion.

- Pre-test probability = 0.33. This value is the highest reported estimation of malingering and/or fraud in the population, and the complement of the pre-test probability used for the first case in the Explanatory category. It is likely that this value is significantly overestimated for two reasons; first, it is the highest value reported in the literature, and the majority of malingering estimates are substantially lower. Second, the rate was reported for all patients with a personal injury claim, rather than those who had claims that had progressed to litigation. It is reasonable to expect that some, if not many of the fraud and malingering cases would be identified prior to litigation.
- True positive rate = 0.65. Hayes et al. refer to the NFL football player study by Newman et al. as the source of their conclusion that a concussion was not consistent with the injury mechanism, as they indicate that the HIP_m they determined for the collision was 5.9 kW and that this corresponded with a concussion risk of less than 10% based on the logistic risk curve derived from the NFL football player data. The potential problems with bias and random error inherent with the use of these data have been discussed earlier in this paper. If the effect of random error (disregarding potential bias) is accommodated for by the bracketing of the data with a confidence interval then a HIP_m of 5.9 kW equates to a maximum concussion risk of approximately 0.35, rather than approximately 0.08. Thus the true positive rate is the probability of no injury at 5.9 kW, which is 0.65.
- False positive rate = 0.35. As discussed above, taking into account the random error in the Newman et al. data, the maximum probability that a concussion would occur at a HIP_m of 5.9 is approximately 0.35.

4.5. Limitations of Error Odds analysis

The Error Odds values for the various case studies as presented in Table 2 varied widely (from 0.3 to 9800), and depended primarily on the pre-test probability of the outcome of interest; the higher the pre-test probability the lower the error rate. Several points are worth noting when reviewing these results:

(1) A number of the input values were highly speculative and/or prone to bias and random variation, although when presented with a range of values on those which favored a higher Error Odds (and thus higher validity) value were used;
(2) Data that were not presented by Hayes et al. in their paper may have been present in the original analysis but not included in the case study, such as the HIC values for the hypothetical restrained traffic crash scenario in case 5. Had these data been used they may have altered the results of the Error Odds calculations;
(3) Although some of the input values were highly speculative, this was largely due to difficulties with the applicability of a biomechanical metric to circumstances in which the metric was inapplicable, and not to an inherent fault with the Error Odds methodology. Examples are the use of the HIC in an anticoagulated patient, or "biomechanical" measure of the probability that an epileptic patient fell during a seizure vs. some other cause, or whether a first time trampoline user was likely to have attempted a back flip. Intent, behavior, and pathology are not measured by biomechanical assessment.

Bearing in mind these caveats, the only case study in Forensic Biomechanics presented by Hayes et al. that had an Error Odds of greater than 10 was the first Explanatory case. None of the other case studies even exceeded an error ratio of 3 correct tests to 1 incorrect test. We find that the promise of Forensic Biomechanics as an ultimate test of truth in disputed matters must be tempered by the potential for error, depending upon how the technique is applied. In cases in which there is a less than 50% pre-test probability of a positive test result, for example when an injury that has been diagnosed and causally attributed to an injury mechanism is refuted by a biomechanical analysis that concludes that that the injury was improbable, the technique can be shown to be prone to the highest levels of error. On the other hand, when

the technique is used in instances with $\gg$50% pre-test probability it is more likely to surpass an Error Odds estimate of 10. We suggest that an Error Odds evaluation of forensic biomechanical analyses be incorporated with the methodology as a means of quality control.

5. Summary

The accuracy of Forensic Biomechanics largely depends on the circumstances in which it is applied and the pre-test probability of the condition of interest. The technique results in the most accurate results when used to explain how an injury occurred, vs. when it is used to refute the causal relationship between an injury mechanism and an injury. The results may be mixed when the technique is used to evaluate the probability of an injury outcome for an actual scenario vs. a hypothetical scenario; we suggest that the biomechanical analysis be correlated with observational epidemiologic study that supports the biomechanical conclusions. A large part of the difficulty with the practical application of Forensic Biomechanics stems from the dichotomization of continuous variables such as risk into an "injury likely" vs. "injury unlikely" delineation, as this raises the potential for false positive results, particularly when the technique is used in an exclusionary manner. An Error Odds analysis that takes into account the pre-test probability of the test result, as well as the true and false positive rate of the test is an important tool for evaluating the accuracy of the forensic biomechanical analysis. A forensic biomechanical opinion that is supported by an Error Odds estimate of 10 or more is more likely to survive a Daubert challenge in the courts.

Conflict of Interest

Both authors provide consultation in forensic matters.

References

1. Hayes WC, Erickson MS, Power ED. Forensic injury biomechanics. *Annu Rev Biomed Eng* 2007;**9**:55–86.
2. Augenstein J, Digges K. Performance of advanced air bags based on data William Lehman injury research center and new NASS PSUs. *Annu Proc Assoc Adv Automot Med* 2003;**47**:99–101.
3. Natl Transp Saf Board. Highway special investigation report: bus crashworthiness. NTSB/SIR-99/04, PB98-917006; 1999. pp. 1–152.
4. Massey v. Newett. Superior court, county of Tippecanoe, Indiana, No. 79D01-0211-CT-00119.
5. Robertson L. *Injury epidemiology*. Cary (NC): Oxford University Press; 1992.
6. Kao SY. Carpal tunnel syndrome as an occupational disease. *J Am Board Fam Pract* 2003;**16**(6):533–42.
7. Clarke M. The control of insurance fraud; a comparative view. *Br J Criminol* 1990;**30**:1–23.
8. Tencer AF, Kaufman R, Ryan K, Grossman DC, Henley BM, Mann F, et al. Femur fractures in relatively low speed frontal crashes: the possible role of muscle forces. *Accid Anal Prev* 2002;**34**(1):1–11.
9. Freeman MD, Rossignol AC, Hand ML. Forensic epidemiology: a systematic approach to probabilistic determinations in disputed matters. *J Forensic Leg Med* 2008;**15**(5):281–90.
10. Newman JA, Barr C, Beusenberg M, Fournier E, Shewchenko N, et al. A new biomechanical assessment of mild traumatic brain injury. Part 2 – results and conclusions. SAE Pap. 2000-13-0017; 2000. p. 2–12.
11. Newman JA, Shewchenko N, Welbourne E. A proposed new biomechanical head injury assessment function: the maximum power index. *Stapp Car Crash J* 2000;**44**:1–35. SAE 2000-01-SC16.
12. Eppinger R, Sun E, Bandak F, Haffner M, Khaewpong N, et al. Neck injury criteria. In: Eppinger R, Sun E, Bandak F, Haffner M, Khaewpong N, et al., editors. *Development of improved injury criteria for the assessment of advanced automotive restraint systems – II*. Washington (DC): NHTSA; 1999. p. 1–23.
13. Mertz HJ, Driscoll GD, Lenox JB, Nyquist GW, Weber DA. Responses of animals exposed to deployment of various passenger inflatable restraint system concepts for a variety of collision severities and animal positions. In: *Proc Ninth Int Tech Conf Exp Safety Veh, SAE 82047*; 1982. p. 352–68.
14. Prasad P, Daniel RP. A biomechanical analysis of head, neck, and torso injuries to child surrogates due to sudden torso acceleration. SAE Paper No. 841656; 1984.
15. Daubert v. Merrell Dow Pharmaceuticals. (92–102), 509 US 579; 1993.
16. Schuhow v. Metlife. District Ct for Scott County, Iowa, No. 98881; 1998.
17. Bridges v. Cates. 401st Judicial District Collin County, Texas, No. 401-00679-04; 2004.
18. Larson v. Bohrer. 7th Judicial Circuit, County of Pennington, South Dakota, No. CIV99-421; 1999.
19. Contreras v. University Neurosurgical Associates. County of Berrien, Michigan, No. 02-3839-NF-M; 2002.
20. Gibson v. State Farm, Civil District Court for the Parish of Orleans, Louisiana, No. 99-02970.
21. Corapi v. Spontak. 2005. 18th Judicial District Brevard County, Florida, No. 05-2001-CA-5915-XXXX.
22. Thomas v. Reed. Circuit Court of the City of St. Louis, Missouri, No. 022-11510; 2002.
23. Bonilla v. New York City Transit Authority. New York, No. 295 AD2d297; 1997.
24. Azzano v. O'Malley-Clements. No. 710 NE, 2D 373.
25. Suanez v. Egeland. 353 NJ Super, No. 191 801A2D 1186.
26. Shultz v. Wells. Colorado Appellate Court, Colorado, No. 99CA0688; 1999.
27. Paulos JA. *Innumeracy: mathematical illiteracy and its consequences*. New York: Hill and Wang; 1988.
28. Thompson WC, Schumann EL. Interpretation of statistical evidence in criminal trials; the prosecutor's fallacy and the defense attorney's fallacy. *Law Hum Behav* 1987;**11**(3):167–87.
29. Leung WC. The prosecutor's fallacy – a pitfall in interpreting probabilities in forensic evidence. *Med Sci Law* 2002;**42**(1):44–50.
30. Bayes T. An essay towards solving a problem in the doctrine of chance. *Philos Trans R Soc* 1763;**53**:370–418. In: Barnard GA. Studies in the history of probability and statistics: IX. *Biometrika* 1958;**45**:293–315.
31. Dawid AP, Mortera J. Forensic identification with imperfect evidence. *Biometrika* 1998;**85**:835–49. Correction In: *Biometrika* 1999;**86**:974.
32. Dawid AP. Statistics on trial. *Significance* 2005;**2**:6–8.
33. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The evidence-based medicine working group. *JAMA* 1994;**271**(9):703–7.
34. Freeman MD, Rossignol AC, Hand ML. Applied forensic epidemiology: the Bayesian evaluation of forensic evidence in a vehicular homicide investigation. *J Forensic Leg Med* 2009;**16**(2):83–92.
35. Report of the Body Engineering and Automotive Safety Committees. Human tolerance to impact conditions as related to motor vehicle design. SAE J885; 2003.
36. <http://www.cdc.gov/ncipc/wisqars/nonfatal/quickpicks/quickpicks_2007/unintall.htm>.
37. Cohen TH, Smith SK. *Civil trial cases and verdicts in large counties, 2001*. US Dept of Justice, Bureau of Justice Statistics NCJ 202803; 2004.
38. Trueblood W, Schmidt M. Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *J Clin Exp Neuropsychol* 1993;**15**(4):578–90.
39. Slick DJ, Tan JE, Strauss EH, Hultsch DF. Detecting malingering: a survey of experts' practices. *Arch Clin Neuropsychol* 2004;**19**(4):465–73.
40. Rosenfeld B, Sands SA, Van Gorp WG. Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Arch Clin Neuropsychol* 2000;**15**(4):349–59.
41. Frederick R, Sargaty S, Johnston JD, et al. Validation of a detour response bias on a forced-choice test of non-verbal ability. *Neuropsychology* 1994;**8**(1):118–25.
42. Rogers R, Harrell EH, Liff CD. Feigning neuropsychological impairment: a critical review of methodological and clinical considerations. *Clin Psychol Rev* 1993;**13**(3):255–74.
43. Insurance Research Council (IRC). Fraud and buildup in auto injury insurance claims, Malvern, PA; 2004.
44. Boström O, Bohman K, Haland Y, Kullgren A, Krafft M. *New AIS1 longterm neck injury criteria candidates based on real frontal crash analysis*. Montpellier, France: Int Res Counc Biomech Impact; 2000.
45. Smeesters C, Hayes W, McMahon T. Disturbance type and gait speed affect fall direction and impact location. *J Biomech* 2001;**34**:309–17.
46. Nakken KO, Lossius R. Seizure-related injuries in multi handicapped patients with therapy-resistant epilepsy. *Epilepsia* 1993;**34**:836–40.
47. Desai KB, Ribbans WJ, Taylor GJ. Incidence of five common fracture types in an institutional epileptic population. *Injury* 1996;**27**:97–100.
48. Nysted M, Drogset JO. Trampoline injuries. *Br J Sports Med* 2006;**40**(12):984–7.
49. Hodgson VR, Thomas LM. Breaking strength of the human skull versus impact surface curvature. NHTSA Rep. DOT/HS-801 002; 1973. p. 1–185.
50. Hart RG, Boop BS, Anderson DC. Oral anticoagulants and intracranial hemorrhage; facts and hypotheses. *Stroke* 1995;**26**:1471–7.
51. Itshayek E, Rosenthal G, Fraifeld S, Perez-Sanchez X, Cohen JE, Spektor S. Delayed posttraumatic acute subdural hematoma in elderly patients on anticoagulation. *Neurosurgery* 2006;**58**(5):E851–6.
52. Fukutake T, Mine S, Yamakami I, Yamaura A, Hattori T. Roller coaster headache and subdural hematoma. *Neurology* 2000;**54**(1):264.